

ABSTRACT OF THE DISCLOSURE

There are provided a singing voice-synthesizing method and apparatus which is capable of performing synthesis of natural singing voices close to human singing voices based on performance data being input in real time. Performance data is inputted for each phonetic unit constituting a lyric, to supply phonetic unit information, singing-starting time point information, singing length information, etc. thereof. The singing-starting time point information represents the actual singing-starting time point. Each performance data is inputted in timing earlier than the actual singing-starting time point, and has its phonetic unit information converted to a phonetic unit transition time length. The phonetic unit transition time length is formed by a first phoneme generation time length and a second phoneme generation time length, for a phonetic unit formed by a first phoneme and a second phoneme. By using the phonetic unit transition time, the singing-starting time point information, and the singing length information, the singing-starting time points and singing duration times of the first and second phonemes are determined. The singing-starting time point of a consonant (first phoneme) is set to be earlier than the actual singing-starting time point. The singing-starting time point of a vowel (second phoneme) is made coincident with or earlier or later than the actual singing-starting time point. In the singing voice synthesis, for each phoneme, a singing voice is generated at the determined singing-starting time point and continues to be generated for the determined singing duration time. State transition characteristics and effects characteristics may be controlled according to input control information.